Available online at www.sciencedirect.com

ScienceDirect

Procedia Manufacturing 3 (2015) 5512 – 5518

Procedia MANUFACTURING

6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE 2015

To Start Voting, Say Vote: Establishing a Threshold for Ambient Noise for a Speech Recognition Voting System

France Jackson[a], Amber Solomon[b], Kyla McMullen[a], Juan E. Gilbert[a]*

[a]*University of Florida Computer & Information Science & Engineering Department, P.O. Box 116120, Gainesville 32611, US*
[b] *Clemson University School of Computing , 100 McAdams Hall, Clemson 29634, US*

Abstract

Prime III is a multimodal voting system that allows users to use touch or voice to make selections on their ballot. This paper discusses an experiment that evaluated the system's speech recognition at various levels of background noise. An approach to simulate realistic background noise in a controlled environment is described. This approach helped mimic a voter voting in a precinct. The goal of the experiment was to establish a threshold for when distortion occurs and speech recognition accuracy declines. The signal-to-noise ratios (SNR) between the volumes were recorded and the system's accuracy was tested. The result was a suggested threshold of a SNR equal to 1.44 to attain 90% system accuracy. The next phase of this project is to test the level of system interference from ambient noise in an actual voting precinct.

Keywords: Speech Recognition; Voting System; Ambient Noise;

1. Prime III

Prime III is a multimodal voting system that incorporates universal design principles to create an accessible voting experience. Regardless of an individual's ability – whether blind, deaf, amputee, or fully abled – a user votes on the same machine independently, privately, and securely. Prime III uses physical input, including touching a

* France Jackson. Tel.: +1-803-609-1052; fax: +1-352-392-1220.
E-mail address: france.jackson@ufl.edu

doi:10.1016/j.promfg.2015.07.707

computer screen or speaking into a microphone, to mark and cast a ballot [3, 8]. Because the system uses speech as an input, loud background noise may undesirably impact the technology.

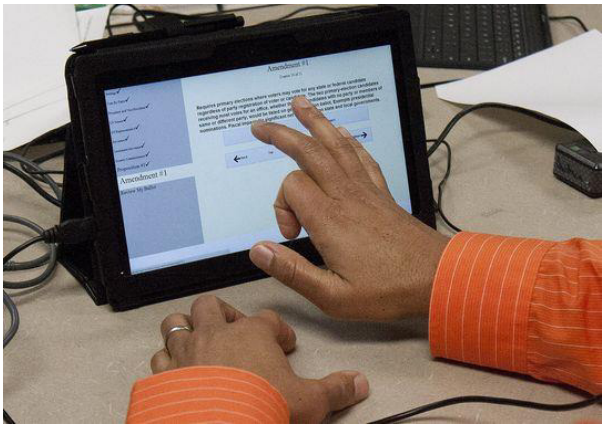

Fig. 1. Prime III is a multimodal system that uses touch and voice for interaction.

2. Introduction

The performance of speech recognition generally degrades in noisy environments or real-world situations [5]. For example, a recognition system with a 1% error rate increased to an error rate of 50% in a cafeteria environment during its busiest hours. Therefore, environmental noise "has become one of the major obstacles to commercial use of speech recognition techniques [2, 5]." This obstacle could be especially problematic for Prime III. Imagine a voter who is blind using Prime III and needing to "speak" a response. In a voting precinct with a high level of environmental noise, Prime III may not recognize the user's selection.

[5, 6] describe two phenomena typically observed in noisy environments that result in accuracy degradation: first, additive noise contaminates the speech signal and, second, speaking in a noisy environment changes data representing speech and causes the Lombard effect. The Lombard effect occurs when speakers attempt to "increase the communication efficiency over a noisy medium [5, 6]."

This paper discusses the effects of varying decibels of background noise on the Prime III Voting Machine's speech recognition. A threshold is provided suggesting when distortion occurs and accuracy degrades with speech recognition.

Fig. 2. A voter uses a headset with an attached microphone to mark their ballot using Prime III.

2.1 How Prime III and Speech Recognition Works

Prime III runs in a web browser, Google Chrome or Firefox, and uses speech recognition to determine if a user made a selection. A headset with a microphone is used to interact with the voting system. Prime III is always detecting sound, because the microphone is always on. However, it only reacts to sounds made during a 1.5 second interval after the user has been prompted. The voter signifies wanting to make a selection by saying, "vote." The following example illustrates the interaction of a voter speaking to mark his or her ballot using Prime III:

Prime III Prompt – "To vote for president say vote"
beep Prime III alerted the voter to speak
…Prime III is now listening for 1.5 seconds
Voter Speaks – "Vote"
…Prime III reacts to the user's selection
Prime III Prompt – "Selected President. You are voting for President. To vote for candidate Gold say vote"
beep Prime III alerted the voter to speak
…Prime III is now listening for 1.5 seconds
Voter says nothing
…Prime III ignores the user's silence
…Prime III goes to the next candidate in the contest
Prime III Prompt – "To vote for candidate Purple say vote"

Prime III determines if the user purposefully responded to a prompt by checking the sound level from the user's microphone against a dynamically determined maximum limit. If the limit is lower than the sound level from the microphone, then Prime III concludes the user did give a response. The value of the limit is determined by the loudness of the background noise. As the level of background noise decreases or increases, the limit decreases or increases, respectively.

2.2 Related Work

There have been multiple studies focused on the issue of speech recognition with varying background noise. [9] developed a recorded test, HINT (hearing in noise test) that measures sentence intelligibility in quiet and in noise.

Another study, performed on the Tangora Speech Recognizer, determined the variance between changes in ambient noises and speech recognition using microphones [2]. This study determined positioning the microphone properly increases recognition performance [2]. Additional experiments have shown "a system trained under a given SNR, signal-to-noise ratio, usually gives poor recognition performance even when tested in a better SNR environment [5, 7]."

[2] focused on microphone characteristics. The experimenters of [2, 5, 7, 9] varied the environments, the microphone settings, and the gender of the speaker's voice.

Although using various locations provides a realistic testing environment, it also provides less control over determining and maintaining background levels. This paper provides an alternative approach that provides a controllable environment that offers realistic background noise.

3. Experimental Work

A total of 108 tests were conducted in an industry standard, carpeted sound booth. Six decibels of background and voice intensities were used: 55, 60, 65, 70, 75, and 80. These decibels were chosen as they represent a typical range of intensities for indoor noise and speech levels [10]. To maintain constant intensities, recordings of the voice and background noises were used. The voter's speech was emulated using an equalized voice recording of a woman in her twenties saying the word "vote." The background noise was simulated using an equalized track from the BBC Sound Effects Library titled "Crowd Milling Around." This track offered high-quality sound with reverberation and natural, realistic fluctuations in intensity [1]. Equalizing both recordings provided a more realistic listening

experience [11].

A background and voice level were paired and tested to determine if interference occurred. For each pairing seventeen scenarios were used to assess the system's ability to understand speech and execute the commands correctly.

1.1 Scenario

The scenarios were designed to test for false positives, to test for recognition of the user purposefully making a selection, and to test detection of the system failing to acknowledge the user making a selection.

False Positive

A false positive occurs when Prime III erroneously deduces the user made a selection. Moreover, Prime III believing the voter said, "vote," when he or she did not. False positives occur for one reason: Prime III believed background noise was the voter speaking. Including scenarios where the voter does not react to a prompt checked for false positives. This provided a way to determine if Prime III mistook background noise for a voter speaking.

Table 1. Frequency response for equipment used

Equipment	Freq. Response
Bose Computer Speakers	65 - 20000 Hz
Logitech Headset	20 - 20000 Hz

Example Scenarios

The following are a few scenarios used describing when Prime III prompted the user, what response the user should give, and the expected response if the system correctly acknowledged the user's selection:

1. System acknowledging command to start voting. This scenario tests for a user purposefully making a selection and if the system acknowledged that selection.
   Example: Prime III- "To start voting say vote"
   Voter–"Vote"
   Prime III- "Selected start voting."
2. System acknowledging command to go to "Settings" Contest. This scenario tests for a user purposefully making a selection and if the system acknowledged that selection.
   Example: Prime III- "To vote for settings say vote"
   Voter–"Vote"
   Prime III- "You are now voting for settings"
3. System acknowledging command to not vote for "Very Fast" candidate in "Settings" Contest. This scenario checks for false positives.
   Example: Prime III- "To vote for very fast say vote"
   Voter– Says Nothing
   Prime III- "To vote for average say vote"

3.2 Materials

The following materials were used for this experiment:

- Industry Standard Sound booth (Double Walled IAC Audiometric Test Room)
- 2 13in MacBook Computers
- 1 27in iMac Computer
- 1 set of Bose Computer MusicMonitor Speakers

- 1 Bose SoundLink Mini Bluetooth Speaker
- 1 set of Logitech Ls11 2.0 Speakers

3.3 Laboratory Setup

The setup was designed to mimic a voter marking their ballot in a voting precinct. There were two computers inside the sound booth: one to use MatLAB to calculate the SNR for each voice intensity and background pair while running Prime III, and another computer to play the voice recording. In order to interact with the computers inside the sound booth without interfering the setup, a wireless mouse and keyboard, and a set of speakers were placed outside the sound booth. The wireless keyboard and mouse, with the speakers, were used to play the voice recording when prompted. The third computer, not inside the sound booth, was used to play the background noise.

Inside the booth two sets of speakers were placed equidistant from one another to create a surround sound effect [11]. The 2 left-hand speakers were matched to the left channel, and the 2 right-hand speakers were matched to the right channel [11]. A single speaker was placed between the earpieces of a headset with a microphone to mimic the voter “speaking” and “listening” to the sounds in the sound booth. The speaker was positioned in the same location as a voter’s mouth to help control variability in speech intensities. The sensitivity of the microphone was set at the standard 50%. A picture of the sound booth setup can be seen in Figure 3. Figure 4 is a diagram of the measurements between each speaker and the placement of each speaker.

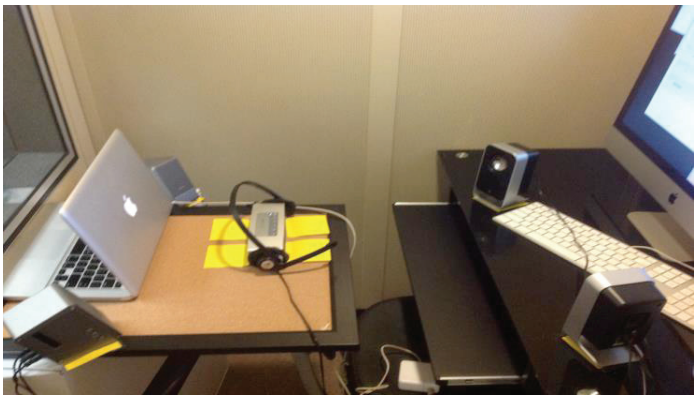

Fig. 3. Computer 1 was used to run the voting system, and Computer 2 was used to play the voice commands. The four outside speakers created the surround sound effect, while the center speaker played the voter’s voice into the microphone.

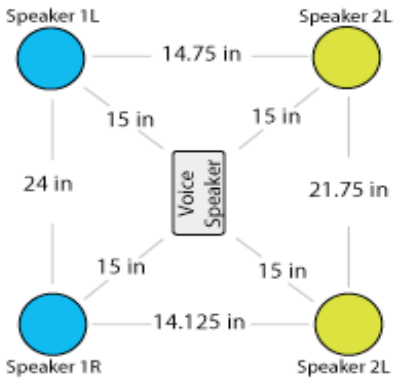

Fig. 4. Measurements between each speaker

3.3 Data Collection

MatLAB was used to calculate the signal-to-noise ratio between the voice and ambient noise and to measure the intensity level. To begin the experiment, an intensity was selected for the background and voice noise from the predetermined range. The volumes of the speakers to play each noise were carefully adjusted until the desired decibels were achieved. There was allowed an approximate 0.50 difference in the actual versus desired decibel value.

Once the conditions were met, the ambient noise recording, with specified intensity, played while Prime III ran. Each scenario was tested and recorded: a failure was marked as 0 and a success was marked as 1. If a scenario permitted a user's response, when Prime III prompted for a reply the voice recording at the specified intensity was played. A ballot with 10 contests was used. The entire ballot was tested at once to determine if intelligibility either worsened or improved as the system progressed. Each background and voice level pairing was tested three times.

4. Results and Discussion

On average Prime III responded correctly to 15 out of the 17 scenarios. Therefore, an accuracy of 90% with a signal-to-noise ratio of at least 1.44 is considered ideal. Speech recognition in an environment with an SNR of less than 1.44 showed a speech recognizer that performed poorly. The intelligibility stayed fairly constant as the ballot progressed.

Regardless of the SNR, there were two main faults observed that often happened. One fault involved testing the first scenario, which was tested as soon as Prime III started. On average, the system was not able to detect that the voter made a selection. The sound recognition did not have ample time to properly adjust its maximum limit to curtail the background noise. The second fault involved Prime III making false positives. This occurred because the sound recognition's maximum limit was too low and natural fluctuations in the background noise were able to surpass that limit. Both, however, were considered system issues and not background interference. Moreover, slight tolerance, 10%, was allowed to account for these issues and any human error that would adversely affect the accuracy.

4.1 Human Error

Human error mainly occurred with playing the voice recording at the perfect time. After Prime III alerts the user to speak, there is only a 1.5 second timeframe for the user to respond. This caused some issues as at times it seemed like the system did not respond, but in actuality the timeframe to respond had closed.

4.2 Voice Recording and Microphone

Using a female voice in their mid-twenties may have advantageously affected the results. Someone older or of a different gender may not have the same voice quality. Words may not be enunciated and articulated as well causing the potential of slight degradation in speech recognition [4].

The microphone settings also may have affected accuracy in speech recognition. The positioning of the microphone was far from the user's mouth. Even with the noise-cancelling property, it still may be difficult for the system to differentiate background noise due to this distance. Prime III uses a specified microphone, so there are no plans to test using different headsets with microphones.

4.3 Outliers

Table 2 highlights some outliers in the results. When the ambient noise was approximately 65 decibels and the voice level was approximately 55 decibels, an SNR equalling 0.1139 below the recommended threshold of 1.44, the system accuracy was over 90%. It is unclear why this occurred. Moreover, when the voice intensity was 55 decibels and the ambient intensity was 60 decibels, the accuracy was only 70%.

Another outlier occurred when the background was approximately 55 decibels and the voice was approximately

60 decibels. The SNR was over 1.44 (2.8228), but the accuracy was just 47%. The cause could be human error, as this was one of the first tests ran, and there was a learning curve involved.

When the background noise is at least 80 decibels the system did not detect any voting attempts. However, the system also did not make any false positives. This was probably due to Prime III determining a maximum limit that was too high that the voter's speech was never able to surpass it.

5. Conclusion

The goals of this experiment were to assess the system's ability to understand and execute voice commands and to establish a threshold for ambient noise to determine when speech recognition degrades. A setup using speakers inside a sound booth was used to imitate a voter marking their ballot using Prime III in a voting precinct. It was determined a SNR of at least 1.44 resulting in a 90% accuracy is ideal. At loud voice intensities where the voice intensity matches the background intensity, accuracy suffers. The SNR needs be above one in order to avoid interference with speech recognition systems.

Although the goal of this study was to improve a specific system, the resulting threshold can be applied to similar speech recognition systems that are used in areas where background noise is present. Most systems that have been tested to date use sound as an optional interface interaction. For some Prime III users, sound is the only method of interaction. Therefore, the implications of this work will be used to improve the universal experience of voters who are normally unable to vote independently.

For this experiment 17 scenarios were tested in a lab setting. The next steps are to test the levels of system interference at actual voting precincts. This would also include testing a wider mix of ages and genders.

Acknowledgements

This material is based upon work supported by the U.S. Election Assistance Commission (EAC). Opinions or points of views expressed in this document are those of the authors and do not necessarily reflect the official position of, or a position that is endorsed, by the EAC or the Federal government. This material is also based in part upon work supported by the National Science Foundation under Grant Number IIS-0738175. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation..

References

[1] BBC Sound Effects Library. Crowd Milling Around. Essential Crowd Sound Effects. (2011). CD.
[2] Das, S., Bakis, R., Nadas, A., Nahamoo, D., and Picheny, M. Influence of background noise and microphone on the performance of the IBM Tangora speech recognition system. IEEE International Conference on Acoust. Speech Signal Process, Vol. II (1993), 27-30.
[3] Dawkins, S., and Gilbert, J.E., Accessible, private, and independent voting. User Experience, Vol. 9, Issue 2 (2010), 16.
[4] Dubno, J. R., Lee, F. S., Matthews, L. J., and Mills, J. H. Age-related and gender-related changes in monaural speech recognition. Journal of Speech, Language, and Hearing Research: JSLHR, Vol. 40, Issue 2 (1997), 444-452.
[5] Gong, Y. Speech recognition in noisy environments: A survey. Speech Communication, Vol. 16, Issue 3 (1995), 261-291.
[6] Juang, B.H. Speech recognition in adverse environments. Computer Speech and Language, Vol. 5, 275-294.
[7] Kitamura, T., Ando, S., and Hayahara, E. Speaker-independent spoken digit recognition in noisy environments using dynamic spectral features and neural networks. Internal. Conf on Speech and Language Processing, Vol. I (1992), 699-702.
[8] Lola, P., Eugene, W., Hall, P., and Gilbert, J.E. Balloting: Speeding up the voting process. HCI International Conference, Vol. 374 (2013), 373-377.
[9] Nilsson, M., Soli, S.D., and Sullivan, J.A. Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. The Journal of the Acoustical Society of America, Vol. 95, No. 2 (1994), 1085-1099.
[10] Noise Sources and Their Effects. https://www.chem.purdue.edu/chemsafety/Training/PPETrain/dblevels.htm.
[11] Standard, E. T. S. I. Speech and multimedia Transmission Quality (STQ); Speech quality performance in the presence of background noise; Part 1: Background noise simulation technique and background noise database. (2012).